Who Gains from Creative Destruction? Evidence from High-Quality Entrepreneurship in the United States

by

Astrid Marinoni
University of Toronto - Rotman School of Management

John Voorheis
U.S. Census Bureau

Abstract

The question of who gains from high-quality entrepreneurship is crucial to understanding whether investments in incubating potentially innovative start-up firms will produce socially beneficial outcomes. We attempt to bring new evidence to this question by combining new aggregate measures of local area income inequality and income mobility with measures of entrepreneurship from Guzman and Stern (2017). Our new aggregate measures are generated by linking American Community Survey data with the universe of IRS 1040 tax returns. In both fixed effects and IV models using a Bartik-style instrument, we find that entrepreneurship increases income inequality. Further, we find that this increase in income inequality arises due to the fact that almost all of the individual gains associated with increased entrepreneurship accrue to the top 10 percent of the income distribution. While we find mixed evidence for small positive effects of entrepreneurship lower on the income distribution, we find little if any evidence that entrepreneurship increases income mobility.

**Keyword*:* entrepreneurship, innovation, income inequality

JEL Classification: L26, D63

Marinoni: University of Toronto - Rotman School of Management, 105 St. George Street, Toronto, ON (Canada); astrid.marinoni@rotman.utoronto.ca
Voorheis: U.S. Census Bureau, Suitland, MD (USA); john.l.voorheis@census.gov

1 Introduction

Entrepreneurial activity is central to the process of economic growth and performance (Schumpeter, 1942; Aghion and Howitt, 1992; Davis and Haltiwanger, 1992). As a consequence, policymakers have often proposed incentives to spur entrepreneurship in an attempt to increase the economic development of their country. In the past decade, millions of dollars have been spent on initiatives meant to nurture local entrepreneurs in the hope of creating the next Silicon Valley, the most famous example of a successful entrepreneurial hub (Saxenian, 1994). However, increasing income inequality in areas with high levels of entrepreneurship suggests that the benefits of entrepreneurship may not be broadly shared. This is related to wider concerns that economic activity within and between regions has become more concentrated, and the distribution of income produced from this activity more unequal. A large literature has provided potential explanations for regional variation in trends in inequality (e.g., Goldin and Katz, 2009; Hémous and Olsen, 2014; Katz and Murphy, 1992; Rosen, 1981). However, there is no consensus on which proposed determinant of inequality is most important.

Recently, the popular press and some practitioners have argued that promoting entrepreneurship and reducing income inequality are mutually exclusive. The rationale stems from the need to provide a sufficient reward for entrepreneurs, given the riskiness of their ventures. Innovative, potentially game-changing ideas might be able to generate enormous financial returns. Successful entrepreneurs will most likely receive a large share of these rents, further pushing out the right tail of income distribution and increasing inequality. An alternate view envisions entrepreneurship as a growth and job creation mechanism that is capable of delivering prosperity and economic welfare to a larger share of the population.

This paper aims to examine the relationship between entrepreneurship and inequality in the United States, with a goal of identifying which of these two views is consistent with our evidence. Ours is the first paper to perform this analysis at a fine geographic scale in the United States and is also the first to analyze the effect of entrepreneurship on income mobility. We introduce and leverage new data in our analysis of the effects of entrepreneurship on income inequality, including a novel measure of the level of entrepreneurship, and new small area estimates of income inequality

and income mobility. We use a measure of entrepreneurship which captures both entrepreneurial quality and quantity (Guzman and Stern, 2017), in contrast to approaches that use the number of businesses to measure entrepreneurship. This measure allows us to differentiate among ventures on the basis of their probability of success, providing a guideline to distinguish entrepreneurs whose venture is a mere substitute for employment income from entrepreneurs whose goal is to generate high economic returns through an IPO or an acquisition. Being able to draw this distinction is particularly important when examining income inequality, given the differential economic impact that these ventures might have on the income distribution.

An additional contribution of the paper is to construct a new dataset of income inequality measures at sub-national geographies. Data on income distributions at geographies below the state level has been scarce or non-existent, making studies of the causes and effects of county- or commuting zone-level inequality impossible. We address this gap by constructing commuting zone level income inequality measures calculated from the American Community Survey and IRS 1040 microdata over the period 1998-2015. This rich microdata allows us to create inequality measures at the commuting zone level, which we can then use to examine the impact of entrepreneurship on the whole income distribution.

With these new data sources in hand, our baseline approach to estimating the effect of entrepreneurship on inequality relies on a standard two-way fixed effects specification. It is possible, however, that the endogenous sorting of individuals across commuting zones over time might bias these estimates. To address this endogenous sorting, we develop a new Bartik-style (Bartik, 1991) instrument for entrepreneurship, which is constructed by interacting the cross-sectional geographic distribution of entrepreneurship ten years before our sample period with national-level trends in entrepreneurship.

In general, we find that entrepreneurship has a large and statistically significant effect on income inequality in both fixed effects and IV specifications. In particular, our main IV specification suggests that a one percent increase in entrepreneurship (which roughly corresponds to the establishment of 15 high-growth start-ups) increases the Gini coefficient by 0.00053[1]. We further examine the impact

---

[1]For this calculation, we consider to be a high-growth start-up a semiconductor firm incorporated in Delaware with a patent and a trademark.

of entrepreneurship across the income distribution by estimating the effect of entrepreneurship on ordinates of the Lorenz curve, or equivalently, on top income shares. These results indicate that entrepreneurship increases the share of incomes at the top of the income distribution. In particular, an increase of 1 percent in entrepreneurship corresponds to an increase of the top 1 percent's share of income of 0.08 percentage points. Although the increase in the top 1 percent share could be consistent with broadly rising incomes, we find little evidence of an economically relevant increase of incomes for people in the bottom half of the distribution. This stands in contrast to the top of the distribution (top decile and top 1 percent) where the increase is positive and statistically and economically significant. It is worth also noting that the gain experienced by the top decile appears somewhat negligible when compared to the one experienced by the top 1 percent. Further analyses exploiting the presence of different types of filings in individual IRS returns attempt to shed more light on which category of income earners (i.e., entrepreneurs, investors, employees) are benefiting from high-quality entrepreneurship. Despite the difficulty in unambiguously identifying these individuals, our results suggest that entrepreneurs, shareholders receiving some dividends from C-corporations and partners of S-corporation belonging to the top 1 percent gain the most from an increase in entrepreneurship.

In addition to considering the relationship between entrepreneurship and inequality, we also examine the impact of entrepreneurship on intragenerational income mobility. This allows us to determine whether the increase of inequality we find is due to re-ranking (the movement of people from a lower to a higher income ranking) or if it can be simply attributed to high income ranking people improving their already favorable position in the distribution. In order to measure income mobility, we construct several measures using linked administrative data, including an absolute upward mobility measure similar to Chetty et al. (2014) and income mobility profiles (Van Kerm, 2009). We find that entrepreneurship has no clear effect on income mobility across the income distribution. This weak evidence on any effect of entrepreneurship on mobility suggests that the benefits of entrepreneurship immediately flow to the top of the income distribution, with no "trickle-down" effect on lower income individuals even five years later.

This paper relates to previous work on inequality and growth by contributing to the extremely sparse literature on the effects of entrepreneurship on inequality. In particular, we extend a newly

emerging literature that tries to shed light on the relationship between corporate demography and inequality (e.g., Sørensen, 2007; Cobb and Stevens, 2017). Within this stream and of literature and closely related to our paper, the work of Sorensen and Sorenson (2007) focuses on the relationship between the number of established businesses and the income distribution in Denmark. In general, the authors find that an increase in the number of businesses tends to worsen income inequality. However, they take into consideration the whole universe of entrepreneurial firms, not distinguishing on their inherent quality. A step in this direction is instead provided by the work of Halvarsson et al. (2018) which distinguishes self-employed and incorporated self-employed entrepreneurs. Both types of entrepreneurship are correlated with an increase in income inequality, although in different sections of the income distribution. While the current results on the relationship between entrepreneurship and income inequality are suggestive, there is a lack of causal evidence, as existing estimates do not attempt to address endogeneity and reverse causality. This paper tries to shed more light on this issue by leveraging new data on entrepreneurial quality, income inequality, and mobility at a micro-geographical level. We also employ an instrument-based identification strategy that might offer more insight into the causal relationship between the phenomena of interest.

Another strand of literature investigates instead the effect of innovation on inequality using patents data. For instance, Aghion et al. (2015) find a positive relationship: innovative states also have high inequality. Similarly, Paunov and Guellec (2017) find that the top income earners have benefited from digital innovations, defined as new products and processes based on software code and data. Other works link innovation to income segregation and come to a similar conclusion (Berkes and Gaetani, 2019). While these results are extremely interesting, they mostly consider innovation and not entrepreneurship. Although one may argue that patenting is fundamental in the appropriation strategy of a new start-up, we must keep in mind that the large majority of patents are filed by large, established firms (Acs and Audretsch, 1988; Cohen et al., 2000).

Finally, our paper extends a vast literature that has examined the potential causes of increasing income inequality in the US (e.g., Goldin and Katz, 2009; Hémous and Olsen, 2014; Katz and Murphy, 1992; Rosen, 1981). This literature has proposed a number of potential explanations – including skill-biased technological change (Card and DiNardo, 2002), changes in the labor market including de-unionization and changes in the minimum wage (Autor et al., 2016; DiNardo et al.,

1996), or changes to the rate of return on capital and the relative importance of capital or business income (Piketty and Goldhammer, 2014; Smith et al., 2019). Our results most closely align with this final strain of literature – we find that high quality entrepreneurship's effects are strongest for individuals with capital and/or business income.

The remainder of the paper is organized as follows: section 2 offers a simple conceptual framework on the possible effects of entrepreneurship on inequality, while section 3 introduces our data. Section 4 presents the basic empirical framework and description of our identification strategy. Section 5 discusses the effect of entrepreneurship and income inequality, while sections 6 and 7 presents results linking entrepreneurship with average incomes and income mobility, respectively. Section 8 concludes.

2 Entrepreneurship and Inequality

The mechanisms underlying the connection between entrepreneurship and income inequality are potentially difficult to tease out. On the one hand, the notion of entrepreneurship is often associated with widespread economic growth. For instance, a vast literature has documented a connection between entrepreneurship and job creation: start-ups account for 20 percent of firm-level gross job creation each year on average, while high growth start-ups have been found to contribute to this figure up to 50 percent (Decker et al., 2014). Agglomeration economies also suggest positive aspects of business creation. Generally, high levels of entrepreneurship in a given area are associated with thicker labor markets, knowledge spillovers, and increases in overall productivity (Marshall, 1920). This could potentially translate into more competitive wages for the employees of successful start-ups. On the other hand, it is also possible that entrepreneurship accrues benefits to risk-tolerant founders and investors providing above-market financial returns. The incomes of these founders and investors might diverge from the population mean (conditional on the success of their venture), thus exacerbating inequality. This view is akin to the winner-take-all hypothesis, by which the most successful competitors are able to capture a disproportionate share of economic rewards (Frank and Cook, 1996). Overall, research has yet to determine which effect in this trade-off is dominant.

To complicate matters further, the analysis would ideally consider the financial position of the entrepreneurs right before the establishment of their entrepreneurial venture. It might be possible that, if most of the individuals have a relatively low income before becoming founders, then entrepreneurship could be a source of economic mobility that might decrease or have little effect on inequality. On the other hand, if the start-ups' founders are from affluent backgrounds, their entrepreneurial effort might compounds gains at the top of the distribution. This latest view is consistent with a stream of literature which views entrepreneurship as a perpetuating reproduction mechanism of social and economic inequalities (Keister and Moller, 2000). In this sense, inequality might be hardwired into a social structure and can inhibit the chances for some people to start a high growth business in the first place.

It is also worth considering which economic agents will be mostly influenced by successful entrepreneurial ventures. Besides the entrepreneur or the founding team, early investors, and key employees of the venture might experience a substantial economic return. In general, however, it is very difficult to quantify returns for employees given due to the presence of different employee equity ownership plans and lockup agreement that might limit the ability of employees to cash out. It is also difficult to identify founding team members without very high temporal resolution data linking employees to employers.

The empirical evidence on entrepreneurship and inequality is scarce in general. Among the related analyses, there is some evidence suggesting a positive relationship between innovation and inequality. In a recent study, Aghion et al. (2015) examine the relationship between innovation and several measures of inequality in the US, finding that states with higher quality-adjusted patents tend to have higher top income shares but also higher social mobility. They also argue this relationship is causal, with an increase in patents per capita explaining 17 percent of the total increase in the top 1 percent income share between 1975 and 2010. Other works also highlighted the existence of a positive relationship between innovation and income segregation (e.g., Berkes and Gaetani, 2019).

Other studies have explicitly focused on entrepreneurship. Sorensen and Sorenson (2007) have examined new firm start-ups in Denmark, finding that an increase in entrepreneurial activity, measured as the number of newly-established firms, is associated with an increase in wage dispersion.

The main mechanism underlying this result is that an increase in the number of firms in a given industry pushes them to differentiate themselves from one another. This, in turn, widens the gap within wages, since discrimination among workers based on their skills and talent allow the presence of different wages for the same job. Despite the advantages of their linked employer-employee data, it is important to note that the way entrepreneurship is measured (i.e., number of establishments) is somewhat limited, as it does not provide any information on the type or quality of ventures. Entrepreneurs seem to be very heterogeneous in terms of the ambition and potential of their ventures, so it may not be ideal to treat them as a homogeneous group (Hurst and Pugsley, 2011). A substantial difference exists between the small number of entrepreneurs whose ambition and capabilities are aligned with scaling up a growing business and the more common incidence of entrepreneurs whose activities are mere substitutes for low-wage employment (Schoar, 2010). Most importantly, these two types of entrepreneurship might have substantially different effect on income inequality. Halvarsson et al. (2018) attempt to shed more light on this distinction by separating self-employed and incorporated self-employed entrepreneurs. They find that both types of entrepreneurship are correlated with an increase in income inequality, although in different sections of the income distribution. In this paper, in addition to devising a strategy to address the endogeneity problems, we use a specific, continuous measure that allows scoring start-ups on their inherent quality according to a variety of characteristics.

This is empirically possible thanks to the work of Guzman and Stern (2017). In their recent study, they propose an index that allows researchers to effectively discriminate between "necessity" entrepreneurs (whose activities are a mere substitute of low-wage employment) and "high-quality" entrepreneurs (whose businesses aim to generate above-average returns). It is easier to appreciate the difference between the number of establishments as defined in the County Business Patterns (CBP) data and the Regional Entrepreneurship Cohort Potential Index (RECPI) by mapping the two measures for the states for which the RECPI measure is available. Figure 1 maps the growth of both variables at commuting zone (CZ) level for the period 1988-2014 split by quintiles. While some areas exhibit a general correspondence, some divergent patterns can be observed. Quite surprisingly, the number of establishments seem to experience a substantial growth in states such as Utah, Colorado, and Wyoming, while the RECPI index suggests that only a subset of commuting

7

zones in this states have experienced a growth in entrepreneurial quality. Another striking difference emerges when examining well-known entrepreneurial hotbeds such as California or New York. While the commuting zones belonging to these states do not experience a large growth in terms of the number of firms, their growth in terms of entrepreneurial quality is remarkable.

The index is intended to capture the portion of entrepreneurship associated with high growth and successful financial exit. Generally, these firms tend to be innovative, venture-backed start-ups in dynamic sectors, such as high-tech. In order to shed light on the validity of this measure of entrepreneurship, we plot the RECPI index against the amount of total venture capital investments (in dollars) in the US as well as the amount of investment by deal in the period 1988-2014. Insofar as high-quality start-ups are more likely to attract external financing, given their higher potential to deliver substantial returns, these two measures should be a good proxy for assessing the presence of potentially high growth firms. Overall, Figure 2 demonstrates that this variable tracks venture capital activity during the time span. In particular, the RECPI index appears to be very responsive to the dot-com bubble and the subsequent financial crisis, implying a strong link with the availability of financing opportunities for entrepreneurship.

3   Data

3.1   Entrepreneurial Quality

Data on entrepreneurship, as noted above, are from Guzman and Stern (2017). They measure entrepreneurial quality by linking the probability of a growth outcome as a function of start-ups characteristics observable at or near the time of initial business registration. In particular, the data take into account whether the firm is named after the entrepreneur, whether it is organized in order to facilitate equity financing (e.g., registering as a corporation), or whether the firm develops innovations or trademarks. The entrepreneurship dataset is constructed as follows. First, the relationship between start-up features at founding and successful growth outcomes (defined as achieving an IPO or a significant acquisition) is estimated. Predictions from these estimates are used to generate an average quality score for each zip code in the US for the 1988-2014 period. By

8

multiplying this index for the number of businesses established in a given area-year combination, it is easy to obtain an overall score (the RECPI index) that multiplies average quality for the number of companies founded in that region.

This measure – RECPI – allows to effectively capture the potential of an entrepreneurial venture in a given space and time at the moment of founding, providing a powerful tool for identifying firms that might have a large positive impact on their community while perhaps commanding superior returns for their founders on average. It is important to note that while so far research has focused on the relationship between RECPI and successful entrepreneurial exits, it is also the case the index is able to predict future GDP growth. This points to the fact that this measure might be able to capture an even wider spectrum of successful firms. Due to data restrictions, RECPI is available for 48 states. While the original data are available at the zip code level, the analysis in this paper aggregates these data to the commuting zone level.[2] In particular, we first aggregate zip code level RECPI to the county level and then at the commuting zone level. Given that commuting zones might span different states, we discard some commuting zones that contain states for which one or more underlying zip code level entrepreneurial variable is not available.

3.2 Income Inequality and Mobility

In order to proceed with our analysis, we require data on the income distribution within commuting zones coinciding with the time period for which high-quality data on entrepreneurship are available. However, no suitable official estimates of income inequality at sub-state geographies exist. To address this, we produce a set of new small area income inequality and mobility measures using confidential survey and administrative microdata housed at the US Census Bureau. Specifically, we produce a variety of commuting zone level income inequality measures using American Community Survey (ACS) and IRS 1040 microdata. These two data sources each have strengths and weaknesses, covering different target populations, income concepts, and income receiving units. We describe each data source in turn, as well a data linkage strategy which can be used to combine these two data sources to create hybrid measures of household income inequality

---

[2] Using David Dorn's crosswalk: https://www.ddorn.net/data.htm

The American Community Survey is a large, nationally representative sample of the entire US population. Since 2005, the ACS has sampled approximately 1 percent of households (about 2 million households) each year, which amounts to about 4 to 5 million annual individual responses. This very large sample size means that aggregate statistics for sub-national geographies (like commuting zones) are reliable for all but the lowest population jurisdictions. The ACS contains information about both household structure and income receipt, which are useful for measuring income inequality. We use ACS households (defined as all individuals, related or not, who reside at a given address) as an income receiving unit and the ACS definition of total household income as our income concept for the ACS-derived measures. ACS total household income includes all income received by any household member from wages, self-employment, investments, and transfer income from Social Security, Supplemental Security Income and other cash benefit programs such as Temporary Assistance to Needy Families (TANF).

The second source of income microdata we use to calculate measures of CZ level income inequality comes from the universe of IRS 1040 tax returns. The Census Bureau is authorized by IRS to use an extract of these universe files, containing information on tax unit characteristics, geography and limited information on income received by tax units, to perform research which benefits Census operations and production activities. These universe files are available for a longer period of time – 1998 through 2015 – and are larger than the ACS. The IRS 1040 universe is the full population of tax filers, and thus represents a different population than the population targeted by the ACS sample. Additionally, the income receiving unit – tax units, which include all primary filers and their dependents listed on form 1040 – and income concept – adjusted gross income, which is all pre-tax income less above-the-line deductions – are substantially different from the ACS. The particular case where the ACS and IRS will differ on income receiving unit definitions involves non-married couples: cohabiting unmarried couples would be in a single ACS household but would comprise two IRS tax units. The differences between the IRS 1040 income concept and the ACS income concept primarily come down to realized capital gains, which are reported on Form 1040, but not in the ACS.

We calculate several measures of income inequality for commuting zones as defined by USDA,

using the 2000 boundary definitions.[3] The most familiar of these is the well-known Gini coefficient, calculated as:

$$G = \frac{1}{\mu_x 2n^2} \sum_{i=1}^{n} \sum_{i'=1}^{n} |x_i - x_{i'}|$$

The Gini coefficient can also be defined as twice the area between the Lorenz curve and the 45-degree line in a standard Lorenz curve graph. The Gini coefficient is a commonly used measure of overall income inequality, but, in our setting, does not necessarily inform which parts of the income distribution are most affected by a particular factor (such as the level of entrepreneurship in a CZ). One way to examine this is to directly use the ordinates of the Lorenz curve as measures of income inequality. The pth ordinate of a CZ's Lorenz curve is:

$$L(p) = \frac{1}{\mu} \int_{0}^{F^{-1}(p)} x f(x) \, dx$$

which we estimate using the Binder and Kovacevic (1995) estimator. These Lorenz curve ordinates can be interpreted as the share of income accruing to the bottom pth percent of the income distribution, and so $1 - L(p)$ can be interpreted as top income shares a la Piketty and Saez (2003). Note that both the Gini coefficient and the Lorenz curve ordinates capture a relative inequality concept. We may additionally be interested in absolute inequality (e.g., how much income (in dollars) goes to the top 10 percent versus the bottom 10 percent), which we capture by reporting the average income within decile bins. For each inequality measure, we adjust for income receiving unit size using a square root equivalence scale. That is, we use $x = \frac{y}{\sqrt{n}}$ in the above formulas, where y is the nominal household (tax unit) income, and n is the number of members in the household (tax unit).

Figure 3 maps the bivariate relationship between entrepreneurship and Gini, using the IRS data. In particular, it considers different percentiles of both measures in the latest available year (2014). Dark blue color denotes those commuting zones which have a substantially high value (in the top third of both distributions) of entrepreneurship and inequality. Perhaps not surprisingly, commuting zones in California, part of the Eastern Seaboard, and Texas are a few of the numerous localities driving the relationship. Overall, only some areas in the central part of the US and Alaska present

---

[3]See here for more details: https://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas.

a high level of inequality and low values of entrepreneurship, while high values of entrepreneurship and low values of inequality are instead more scattered.

Next, we plot the values of entrepreneurship and inequality by year to examine the naive correlation over time and space. Figure 4 reports a scatterplot mapping RECPI and inequality variables for all the variables for selected years. A superficial visual inspection suggests a positive relationship between the two variables. Again, California plays a dominant role: the Los Angeles and San Francisco Commuting Zones display one of the highest RECPI and inequality values. This trend is quite consistent across the years. As noted in the previous figure, some areas in Wyoming display instead a very high inequality across all years, but very low levels of entrepreneurship. Overall, it seems that the relationship is positive and it is not particularly sensitive to outliers. A similar plot can be obtained even excluding all the commuting zones in California.

4 Empirical Strategy

Our goal is to identify the effect of commuting zone level entrepreneurship, measured by the RECPI index, on the distribution of income within commuting zones, captured by the income inequality measures described above. As we have both cross-sectional variation across commuting zones and temporal variation within commuting zones over time, a simple approach to estimate the effect of entrepreneurship on income inequality would be to estimate two-way fixed effects regressions of the form:

$$Ineq_{i,t} = \alpha + \beta RECPI_{i,t} + X_{i,t}\Gamma + \lambda_i + \delta_t + e_{i,t} \tag{1}$$

where i indexes commuting zones and t indexes year. $RECPI_{i,t}$ is our measure of entrepreneurship. λ_i and δ_t are commuting zone and time effects respectively. Give the RECPI index is available from 1988 to 2014 we are able to run the above regression from 1998 to 2014 for inequality measures based on IRS data and from 2005 to 2014 for inequality measures using ACS data. In all regressions, we cluster standard errors at the commuting zone level. The interpretation of β is the causal effect of entrepreneurship on inequality in our regression amounts to an assumption of selection-on-observables: $RECPI$ should be independent of income inequality, conditional on time-invariant

CZ characteristics, year fixed effects and observable time-varying CZ characteristics.

The chief threat to identification in this setting arises from unobservable, time-varying characteristics of commuting zones that drive the RECPI-inequality relationship. Specifically, it is possible that locational sorting induces this type of endogeneity. If low and high income individuals sort differentially into high or low entrepreneurship areas, this could result in a correlation between RECPI and income inequality as this locational sorting affects the CZ income distribution. Note that the direction of the endogeneity bias due to this sorting is uncertain a priori. If high income individuals sort from low RECPI CZs to high RECPI CZs, this would increase income inequality in high RECPI areas and bias β upward. If instead low income individuals sort from high RECPI areas to low RECPI areas, this would result in an increase in inequality in low RECPI areas and a downward bias of β. Thus, the sign depends on the relative magnitude of low and high income cross-CZ migration flows. In the face of this endogeneity, it is necessary to find some source of identifying variation that can be leveraged, such as an instrumental variables setting. We can estimate two-stage least squares regressions of the form:

$$
\begin{cases}
RECPI_{i,t} & = \gamma + \theta Z_{i,t} + X_{i,t}\Lambda + \lambda_i + \delta_t + \epsilon_{i,t} \\
Ineq_{i,t} & = \alpha + \beta_{IV} \widehat{RECPI}_{i,t} + X_{i,t}\Gamma + \lambda_i + \delta_t + e_{i,t}
\end{cases}
\tag{2}
$$

Where $Z_{i,t}$ is our proposed instrument, i.e., a Bartik-style (shift-share) instrument for entrepreneurship.

This Bartik-style instrument is constructed by interacting the cross-sectional distribution of entrepreneurship within CZ before the estimating sample with national level trends in entrepreneurship. Specifically, we begin by fixing the cross-zip code distribution of RECPI levels in 1988. We then assign each zip code to the decile of the national zip-code level RECPI distribution. We calculate each zip code's predicted RECPI by allowing each zip code grow at its decile's national level growth rate over the period 1989-2014, and then calculate CZ-level predicted RECPI levels by summing across these predicted zip code levels (weighted by allocation zip code to CZ factors). This predicted CZ RECPI then is used as an instrument for RECPI. Its construction is designed to address potential endogeneity by fixing the distribution of RECPI in 1988.

This class of Bartik-style instruments has become an increasingly common tool in applied empirical work. The source of identifying variation underlying this identification strategy comes from the initial cross-sectional distribution of entrepreneurship within commuting zones, and thus the (untestable) identifying assumption is that the initial cross-sectional distribution of entrepreneurship is uncorrelated with subsequent unobserved shocks to income inequality.

5 Entrepreneurship and Income Inequality

In this section, we present the results from some fixed effects and instrumented regressions of entrepreneurship on different measures of income inequality. We then look specifically at the relationship between entrepreneurship and the shares of income across the whole distribution as measured by the Lorenz Curve.

5.1 Fixed Effects Regressions

We begin by providing some baseline results from fixed effects models on the correlation between entrepreneurship and selected inequality variables. Table 1 presents the results from the regression of the Gini index calculated with IRS and ACS data on entrepreneurship. Column (1) suggests a positive relationship between RECPI and the Gini index (IRS). This correlation holds if we add commuting zone specific time trends as additional controls in column (2). If we consider the Gini index constructed with ACS data, the relationship is statistically insignificant in both specifications (columns (3) and (4) respectively) but positive. This may be due to a loss of power, given that the period for which ACS data are available (2004-2014) is substantially shorter. It may also be due to the different type of income captured by ACS, which does not include capital gains.

We proceed by assessing whether entrepreneurship has different effects on various points on the income distribution. To do so, we examine the effect on the Lorenz Curve and we use Lorenz Curve ordinates as dependent variables in our regressions. Recall that Lorenz curve ordinates represent the share of income accruing to the bottom x percent of the income distribution, so a marginal effect b on a Lorenz ordinate can be interpreted as a $-b$ effect on the top $1 - x$ percent's share of income.

14

For simplicity, we report coefficients with their sign reversed, so that they reflect an increase of the income share above a particular percentile. Figure 5 presents the coefficients of fixed effects models in which entrepreneurship is regressed on every fifth percentile of the income distribution. Again, we contrast results using Lorenz curve ordinates calculated from our two data sources and consider models without and with commuting zone-specific trends. When using IRS data (Panel A) it is quite clear that entrepreneurship tends to be associated with an increase in the shares of income at the top of the distribution. In particular, this correlation is positive and significant starting at the 60th Lorenz ordinate (top 40 percent). The result is very similar and robust to the inclusion of commuting zones specific trends. Panel B takes into account the ordinates of the Lorenz Curve, which have been built with ACS data. The trends previously found in Panel A seems to be vaguely confirmed even if the standard errors tend to be much higher and all of the coefficients lose their significance. Models including time trends are very much in line with the ones without trends even though a clearer positive and significant relationship can be observed for some percentiles at the top of the distribution.

5.2 IV Regressions

We might be worried that the positive relationship found in the preceding regressions is biased, given the possible endogeneity of our entrepreneurship measures. We proceed then to instrument the measure for entrepreneurship with a Bartik-style instrument described in the previous section.

Table 2 shows the result of regressions using our Bartik-style instrument for RECPI entrepreneurship. Overall, the main results are confirmed. In particular, when examining the effect of entrepreneurship on the IRS Gini index in column (1), one can notice the coefficients are strongly significant and larger than the ones of Table 1. This specification suggests that an increase of RECPI of 1 percent, roughly equivalent to the establishment of 15 high-quality start-ups,[4] increases the Gini index by 0.00053 points. Put in another way, moving from the 25th percentile of CZ RECPI to the 75th percentile of CZ RECPI would increase the Gini coefficient by about 0.01 points – about 46 percent of the average increase in CZ-level Gini between 1998-2015. The

---

[4]We consider being a high-growth start-up a semiconductor firm incorporated in Delaware with a patent and a trademark. Similar figures can be obtained by considering different industries such as e-commerce.

IV coefficients are substantially larger than our fixed effects estimates, suggesting that if anything, the previous estimates were biased downward. Column (2) considers the same regression using CZ-specific trends. The coefficient is again positive and significant, thus reinforcing our previous result. Column (3) and (4) consider ACS-based Gini measures. The coefficients remain positive and significant in both cases, even if the F-statistics in column (4) casts some doubt on the relevancy of our instrument for this particular specification. However, survey-based data seem to confirm the existence of a positive relationship between entrepreneurship and inequality.

We also estimate IV regressions to study the effect of entrepreneurship on Lorenz curve ordinates. Figure 6 presents the coefficients of IV specifications in which entrepreneurship is regressed on every fifth percentile of the income distribution. Panel A partially confirms the results of figure 5: entrepreneurship has now a positive impact on the share of income for a good part of the income distribution, with the top 1 and 5 percent gaining the most. In particular, an increase of 1 percent in entrepreneurship corresponds to an increase in the top 1 percent's share of income of 0.08 percentage points. Put in another way, moving from the 25th percentile of CZ RECPI to the 75th percentile of CZ RECPI would increase the top 1 percent share by about 0.016 – about 84 percent of the average increase in CZ-level top share between 1998-2015. In general, coefficients are statistically significant when considering percentiles greater than the 40th. The model including trends seems to paint a picture which is more in line with the previous fixed effects results: only people above the 65th percentile of the distribution tend to experience an increase in their income share, while people belonging to the lowest section of the distribution are experiencing a negative effect on their shares of income. The models in Panel B considers, as before, inequality measures using administrative (ACS) data. Despite the loss of relevancy of our instrument in some specifications,[5] the models seem to confirm the previous findings, suggesting once again a positive relationship between entrepreneurship and income shares for a substantial part of the income distribution. Overall, the majority of the models in Figure 5 highlight how entrepreneurship has a positive effect on the share of income of people belonging to the top half of the income distribution. Crucially, however, the gains experienced by the middle part of the distribution are overshadowed by the ones accruing to the very top percentiles, which results in an overall large increase in inequality.

---

[5] Regressions where the F-stat of our instrument are below 10 are highlighted with dashed lines.

6   Entrepreneurship and Average Incomes

While there exists a strong positive relationship between entrepreneurship and inequality, it is essential to assess whether entrepreneurship has an impact on the income levels of individuals across the income distribution. It is possible that a positive relationship might perhaps partially mitigate the negative effect of entrepreneurship: while it would be still responsible for an unequal distortion of relative incomes, it might still be able to increase absolute ones, thus potentially improving the economic conditions of people belonging to a given part of the income distribution.

In order to check whether this is the case, we regress average incomes by decile on entrepreneurship, using the Bartik-style IV as in the inequality IV models. Figure 7 plots the coefficients (in dollars) of model regressing the average income of each decile in the income distribution on entrepreneurship. The average income of the 99th percentile is also reported. Panel A uses IRS-based average incomes, while panel B considers ACS-based figures. The basic IRS-based model suggests a significant increase in the average income of the population belonging to the 99th percentile of the income distribution and more generally to those belonging to the 10th decile. In particular, an increase of one percent in entrepreneurship increases the average income of people belonging to the top 1 percent of the income distribution by roughly $10,000. When considering the top 10 decile, this figure decreases to $1,400. Most of the other deciles (from the 3rd to the 6th) are instead negative and significant, suggesting that entrepreneurship might shrink average incomes of people belonging to the middle of the distribution. Adding time trends confirms the main positive and significant relationship for the very top of the distribution, yielding a very similar coefficient. However, these models hint at a possible positive relationship between all the lower deciles and average income, in contrast with the previous specifications.

The ACS-based models in Panel B again confirm the fact that people belonging to the top 10 and 1 percentiles are gaining the most in terms of increase of average income. When looking at the lower part of the distribution, these models again hint at the fact that there might indeed be a positive and statistically significant relationship for all the other deciles in the distribution. In general, the most crucial results emerging from our model is again the disproportionate gain of the very top section of the distribution: while some models suggest that a surge in entrepreneurship

17

might increase the average incomes of the majority of the distribution, this effect does not appear to be economically significant. Once again, the gains in average income of most of the income distribution are dwarfed by those accruing to the top 1 and 10 percent. In particular, a shocking 84 percent of the total gains derived by an increase of entrepreneurship of 1 percent accrue to the top decile of the income distribution, with the top 1 driving this trend.

Note that although a deviation in qualitative results between the ACS and IRS income data could arise for a number of reasons – including mismeasurement in the survey data, under-coverage of top incomes, or misalignment of income concepts – we see very similar qualitative results for both relative income inequality (Gini and Lorenz ordinates) as well as absolute inequality (average incomes by decile).

6.1 Who Benefits from Entrepreneurial Rents? Entrepreneurship and Average Incomes Split by Category

One question that emerges from the previous analyses concerns the categories of people who benefit the most from the entrepreneurial rents. We might be interested in knowing whether is it entrepreneurs, investors, or employees who are driving the increase in income in the top of the distribution. We exploit the presence of different types of filings in the individual IRS tax returns in order to form broad groups having different combinations of incomes. In practice, it is extremely difficult to draw a clear line among the groups that we might be interested in observing, but we can still exploit the presence of different income profiles to draw some interesting insights.

The main category we are interested in observing is entrepreneurs, who might submit different filings depending on the type of companies they own. Specifically, there are three main business forms that they might choose, which are more or less suitable depending on the entrepreneurs' intentions and expectations about the purpose and growth of their company. One of the simplest ways to form a business is through a sole proprietorship. These companies are usually small in nature, present limited growth prospects, and they do not usually have a large number of employees. Given there is no separation between the assets of the business and those of the owner, their risk profile tends to be extremely low. Differently from this type of businesses, there are other forms that

18

offer the advantage of protecting shareholders' personal assets. For instance, some common forms are S-corporations and C-corporations. S-corporations are considered a pass-through entity from a federal tax point of view, meaning that income and losses are passed directly to the shareholders and they are not taxed at the company level. Despite this advantage, S-corporations are limited in terms of the number and nationality of shareholders and the type of stocks they can offer. Specifically, S-corporations do not allow the creation of distinct classes of owners who are entitled to different dividends or distribution rights. This limits their suitability for businesses needing external capital, such as venture capital funding. These companies are usually mid-market firms operating in skill and labor-intensive sectors (e.g., wholesale, financial intermediation, or other business activities). Nonetheless, they might still achieve significant economic success, and some of their owners have been found to belong to the top 1 percent of the income distribution (Smith et al., 2019). C-corporations allow instead more flexibility when it comes to ownership structure and they are usually the preferred company form for high-growth businesses seeking external funding (Darrow, 1989) or for companies in capital-intensive industries (Smith et al., 2019). Unlike S-corporations however, they are subject to the burden of "double-taxation" given that both the entity itself and the shareholders pay taxes on annual income and dividends or capital gains, respectively.

In our data, we are able to observe the presence of a limited set of schedules that might give us some indication about the type of entrepreneur that we are examining. Specifically, we are able to observe the presence of Schedule E, C, and schedules broadly related to investment income. Schedule E needs to be filed by partners or shareholders of an S-corporation and by people who earn rental income. Schedule C is instead filed by sole proprietorship. Owners of C-corps who receive dividends from their businesses would claim these directly on their 1040 forms. We also observe if they filed a Schedule D (related to capital gains or losses) which might give us more information about individuals' income profiles.

One obstacle in our data is the inability to observe the presence of sub-schedules that would help to construct unequivocal categories of entrepreneurs. For instance, it is somewhat complicated to identify C-corporation owners, as not all C-corporation shareholders receive dividends from their firms. In particular, one can argue that owners of early-stage start-ups might tend to reinvest profits in the firm rather than provide a payout to investors. If the owners decide to receive a payment

19

in salary, they would not report dividends on their 1040 return. Also, it is worth noticing that using the AGI income concept measured in the 1040 data may not fully capture the Haig-Simons income[6] of successful startup owners, which will largely be driven by changes in net worth due to unrealized capital gains of their equity in the startup venture.

We construct the income categories taking into account the main combinations using the presence of schedules we can observe (i.e., Schedule C, Schedule E, investment income) and their absence. Despite the limitations, the existent income categories might still provide us some insights and highlight interesting differences. Figure 8 plots, for each category, the coefficients of an IV regression estimating the effect of entrepreneurship on the average income within decile bins (and the top percentile bin). We keep within-CZ percentiles constant across these income categories. Alternate models using within-group percentiles show broadly qualitatively similar patterns.

Once again, for each figure, we provide a zoomed version considering the lowest deciles, given the difference between their coefficients and the ones of the top distribution is substantial. The first two panels in the first row of Figure 8 consider people who filed a Schedule C or Schedule E and do not have any investment income. Overall, an increase in high-quality entrepreneurship does not seem to have a significant effect on the incomes of the top distribution. These individuals represent most likely unsophisticated entrepreneurs that are not likely to establish high-growth start-ups (i.e., sole proprietors or owner of S-corporations with no investment income). This result changes if we consider individuals with Schedule C or Schedule E that also filed some investment-related schedules (second row of Figure 8). In particular, an increase in entrepreneurship now translates in a significant increase in the income of people who file a Schedule C. We suspect that the people in this category might primarily be C-corporation shareholders who also receive some dividends by the firms. This means that they might be shareholders of more mature start-ups, but they might also be shareholders of different start-ups at different life stages. We obtain similar results when we consider individuals who filed a Schedule E and some investment related-schedules. Overall the top decile and the top 1 percent experience a significant economic benefit from the establishment of high growth start-ups although not as much as Schedule C filers. People belonging to this category might be owners of S-corporations but also partners and shareholders of private equity and

---

[6] Haig-Simons includes personal expenditures plus changes in wealth.

venture capitals, which are mostly organized as S-corporations. The last row in Figure 8 considers instead individuals who only filed investment-related schedules or people who did not file any of the schedules we can observe. These two latter groups are even more difficult to identify. In the first group, we might find people with only some investment income, employees with some investment income, and even C-corporation owners receiving a salary or no income at all from their start-ups but who still receive some investment income from other sources. In the second category, we might instead find unemployed individuals not receiving any type of income, employed individual, and again C-corporation owners receiving a salary or no income from their start-ups.

Overall, notwithstanding the major obstacles we are facing in identifying entrepreneurs, investors, and employees of high-growth start-ups, we can still observe some speculative patterns. Specifically, while we cannot observe significant variations in the gains of the majority of the income distribution across categories, there seem to exist some disparity in the income of people belonging to the top 1 percent in Figure 9.[7] Our supposition is that entrepreneurs or shareholders gaining some dividends from C-corporations and partners of possible private equity firms belonging to the top 1 percent (for their income profile) are those who gain the most from high-quality entrepreneurship. However, our results would also be consistent with a story whereby the benefits of entrepreneurship spillover to the owners of other mature closely-held firms, who are important for understanding trends in top incomes more broadly (Smith et al., 2019).

7 Entrepreneurship and Income Mobility

The process of creative destruction in which entrepreneurship plays a central role is a fundamentally dynamic process. This dynamism, in turn, suggests that entrepreneurship may affect not just the level of income (and the cross-sectional distribution of income), but also the level and distribution of income growth. Thus, we turn our attention to how entrepreneurship affects a second set of outcomes measuring relative and absolute intra-generational income mobility. We define two income mobility concepts which correspond to both relative and absolute ideas of intra-generational

---

[7]Since the tax data only flags the presence of investment or business income, not the amount, it is possible that our defined groupings are more homogeneous in the bottom of the distribution than in the top, leading to less variation across groups in the effects at the bottom of the distribution.

mobility. That is to say, these two concepts correspond to the ideas of "rising up the income ladder" versus being "better off than you were a year ago". Both of these, of course, will vary across the initial income distribution, so we choose measures which trace out income mobility at different parts of the income distribution.

To fix ideas, let b be the base year, while f is the final year, and let Y_b, Y_f be income earned in these two years, so that $p_i = F(Y_i), i \in \{b, f\}$ and $x(p_i) = F^{-1}(Y_i), i \in \{b, f\}$. In general, our mobility measures will take the form $m(Y_b, Y_f | Y_b = x(p_b))$. This $m(.)$ function will, in other words, describe some concept of mobility for individuals who start at percentile p_b of the base year income distribution. The first concept we will consider is measured by:

$$m_r(b, f, p_b) = \int_{z_-}^{z_+} p_f(y) dF(Y_f | y_b = x(p_b))(y)$$

that is to say, the average normalized rank in the final year conditional on starting at normalized rank p_b in the base year. This is a *relative* income mobility measure, in that it only consider individuals' rank in the income distribution, and not their absolute level of income. The second concept is measured by:

$$m_a(b, f) = \int_{z_-}^{z_+} log(y_f) - log(y_b) dF(Y_f | y_f = x(p_b))(y)$$

that is to say, the average individual income growth rate between the base and final year, conditional on starting at the p_bth percentile of the base year income distribution.

We calculate each of these two measures for 2000 vintage commuting zones for two different windows of mobility: mobility over a one year horizon, and mobility over a five year horizon. To do this, we link individuals longitudinally across 1040s, assigning the tax unit AGI, adjusted for inflation, to each individual who appears as a primary or secondary taxpayer in each year. We impose a series of sample restrictions on these longitudinally linked 1040s. We restrict geography only in the base year, so that the mobility measures capture potential geographic mobility in response to local economic conditions. We also restrict the sample to only individuals aged 25-54 in the base year to avoid life cycle effects due to labor force entry and retirement.

We adopt our usual empirical approaches to the inequality results above, i.e., we employ our baseline fixed effect specification where we regress mobility between the base year and final year on the characteristics of the commuting zone in the base year, controlling for CZ and year fixed effects. Thus, we estimate regressions of the form:

$$m_r(t, t+k, p_b)i, t = \alpha + \beta RECPI_{i,t} + X_{i,t}\Gamma + \lambda_i + \delta_t + e_{i,t} \tag{3}$$

where we consider time horizons of $k = 1, 5$ years ahead. This specification is designed to capture whether the level of entrepreneurship in a given CZ at a point in time affects income growth and mobility over the next k years. As before, we also use a Bartik-type instrument to address potential endogeneity due to locational sorting, in which case we estimate regressions of the form:

$$\begin{cases} RECPI_{i,t} & = \gamma + \theta Z_{i,t} + X_{i,t}\Lambda + \lambda_i + \delta_t + \epsilon_{i,t} \\ m_r(t, t+k, p_b)i, t & = \alpha + \beta_{IV}\widehat{RECPI}_{i,t} + X_{i,t}\Gamma + \lambda_i + \delta_t + e_{i,t} \end{cases} \tag{4}$$

We will focus on the results for the predicted rank concept (m_r) of income mobility; results using an income mobility profile concept (m_a) are available in the appendix.

Figure 10 visualizes the results of a series of fixed effects regressions of the form of equation 3, estimated separately for each $p \in \{5, 10, ..., 95, 99\}$. The top panel reports results for a one year ahead window, with the right panel reporting results including CZ-specific trends, and the left panel showing results without trends. The bottom panel reports equivalent results for a five year ahead growth window. When examining the models without trends, results are quite mixed and a clear pattern does not seem to emerge. However, the models with trends hint at a possible downward pattern, suggesting an increasingly negative relationship between entrepreneurship and predicted ranks at different percentiles, especially in the 1-year model.

Figure 11 reports results from IV regressions of the form in equation 4, again visualizing estimates for $p \in \{5, 10, ..., 95, 99\}$. As with the previous figure, short-run (one year ahead) results are reported in the top panel, while long-run (five years ahead) results are reported in the bottom panel. The IV specification does not help us in establishing some clear trends in the results. The one year

models suggest a possible increase of mobility for people who start before the 25th percentile and a decrease in the middle part of the distribution. The 5-years model instead seems to suggest an upward relationship, where people starting at the top of the income distribution experience a significant increase in their upward mobility. However, when looking at short and long-run models with trends, it is almost impossible to pick up a pattern in the results as coefficients are more scattered and rarely significant. These results are broadly consistent with our interpretation of the cross-sectional distributional results above, which is to say that the benefits of entrepreneurship almost solely accrue to already high income individuals.

8   Conclusion

Using new aggregate measures of income inequality and income mobility produced from the universe of tax returns linked with survey data, we have surfaced new evidence on an old question – who benefits from creative destruction, which we capture by new measures of high-quality entrepreneurship? In CZ-by-year fixed effects regressions, and in regressions instrumenting for entrepreneurship using a Bartik-style instrument, we find that, overwhelmingly, the benefits of entrepreneurship accrue to high-income individuals, with no evidence of offsetting growth or mobility effects accruing to the bottom of the distribution. Put starkly, our evidence suggests that if anything, entrepreneurship makes the rich richer and the poor poorer.

Although the use of high-quality administrative data allows us to produce highly detailed aggregates of local income distributions, we are not able to fully tease out the mechanisms by which entrepreneurship affects different groups of individuals – for instance investors, the entrepreneurs themselves, and employees of innovative firms. While we attempt to shed some light on this issue, it is difficult to clearly identify these categories using the tax data available in the Census Bureau's linkage infrastructure. Specifically, many of the high-quality start-ups which are captured in the RECPI measure might be incorporated as C-Corporations. These entrepreneurs are particularly difficult to pinpoint using the available administrative data. Future work to understand the mechanisms by which entrepreneurship increases inequality should focus on incorporating firm and establishment level data (such as the Census Bureau's Longitudinal Business Database) to examine

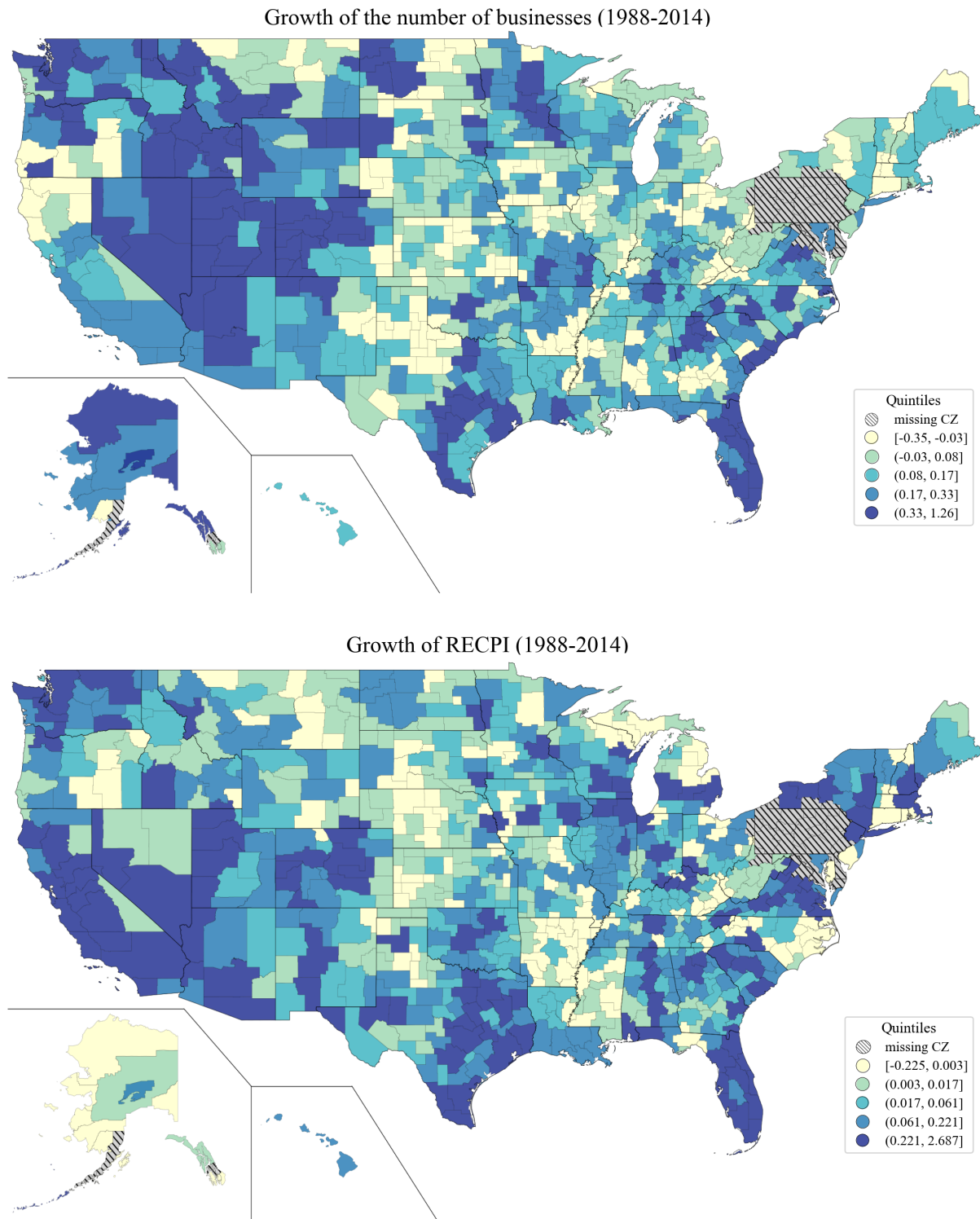how employees of different types of firms are affected.

9 Tables and Figures

Table 1: Fixed effects regressions of entrepreneurship (RECPI index) on the Gini index built from (IRS & ACS data)

	(1)	(2)	(3)	(4)
Model:	FE	FE	FE	FE
Dependent var:	Gini IRS	Gini IRS	Gini ACS	Gini ACS
RECPI	0.010***	0.015***	0.004	0.010
	(0.004)	(0.005)	(0.006)	(0.009)
CZ-Trends	N	Y	N	Y
Observations	11,669	11,669	7,315	7,315
R Squared	0.869	0.901	0.580	0.628

The measure of entrepreneurship is the RECPI index transformed as log(1+RECPI). The dependent variable in columns 1 and 2 is the Gini index calculated with the IRS data (for the period 1998-2014). The dependent variable in columns 3 and 4 is the Gini index calculated with the ACS data (for the period 2004-2014). Columns 2 and 4 include Commuting Zone specific linear trends. All variables are at Commuting Zone level. All regression include controls for the number of unemployed people on the active population, the number of Native American and Alaska Natives, Asian, Black, and Hispanic people on population, the number of people under 25 and over 65 years of age on population, the log number of population, and per capita personal income in dollars. Errors are clustered at the Commuting Zone level: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: Change (1988-2014) in the log number of businesses versus change in RECPI at CZ level (quintiles)

Growth of the number of businesses (1988-2014)

Quintiles
- missing CZ
- [-0.35, -0.03]
- (-0.03, 0.08]
- (0.08, 0.17]
- (0.17, 0.33]
- (0.33, 1.26]

Growth of RECPI (1988-2014)

Quintiles
- missing CZ
- [-0.225, 0.003]
- (0.003, 0.017]
- (0.017, 0.061]
- (0.061, 0.221]
- (0.221, 2.687]

Source: Guzman and Stern (2017) and County Business Patterns. The figure shows the growth of the number of businesses (source: Census County Business Pattern) versus the growth of the RECPI index for the period 1988-2014. Growth values have been calculated as the difference of the log values of the variables in 1988 and 2014. DRB approval numbers CBDRB-FY18-164, CBDRB-FY19-342 and CBDRB-FY19-509.

Table 2: Instrumented fixed effects regressions of entrepreneurship (RECPI index) on Gini index built from IRS & ACS data

	(1)	(2)	(3)	(4)
Model:	IV	IV	IV	IV
Dependent var:	Gini IRS	Gini IRS	Gini ACS	Gini ACS
RECPI	0.053***	0.029***	0.067***	0.059**
	(0.011)	(0.011)	(0.025)	(0.029)
CZ-Trends	N	Y	N	Y
Observations	11,488	11,488	7,203	7,203
R Squared	0.866	0.902	0.578	0.629
F-stat	21.555	23.394	12.080	4.557

The measure of entrepreneurship in all regressions is a Bartik-style instrument of the RECPI index built from 1988 values of RECPI. The dependent variable in columns 1 and 2 is the Gini index calculated with the IRS data (for the period 1998-2014). The dependent variable in columns 3 and 4 is the Gini index calculated with the ACS data (for the period 2004-2014). Columns 2 and 4 include Commuting Zone specific linear trends. All variables are at Commuting Zone level. All regression include controls for the number of unemployed people on the active population, the number of Native American and Alaska Natives, Asian, Black, and Hispanic people on population, the number of people under 25 and over 65 years of age on population, the log number of population, and per capita personal income in dollars. Errors are clustered at the Commuting Zone level: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 2: Panel A: RECPI index and venture capital investments from 1988 to 2014; Panel B: RECPI index and investments by deal from 1988 to 2014.

Panel A

Panel B
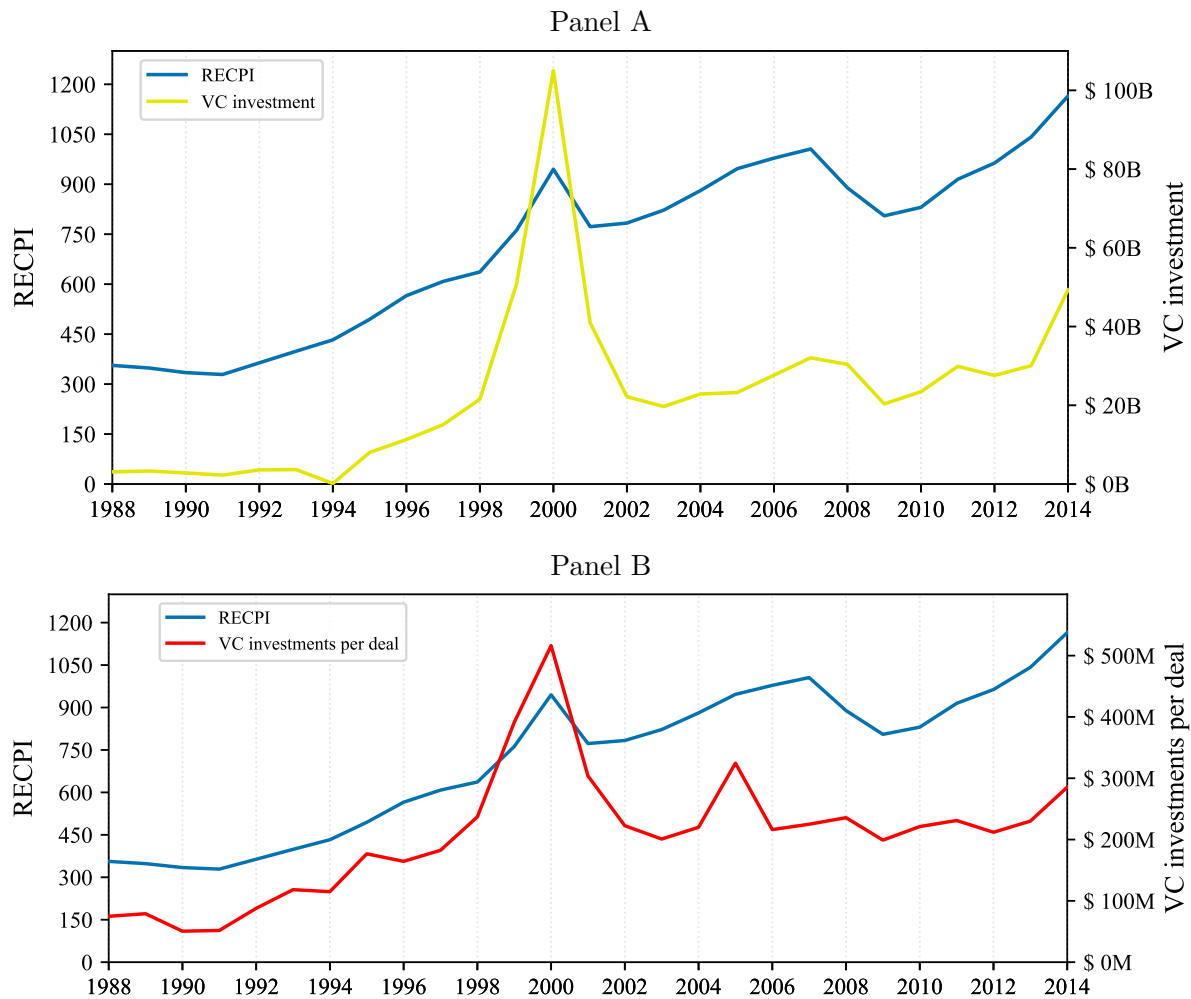
Source: Data on venture capital investments and venture capital investments per deal is provided by the National Venture Capital Association Yearbook 2015-2016 (https://nvca.org/research/research-resources).

Figure 3: Bivariate map of RECPI and Gini (IRS) in 2014
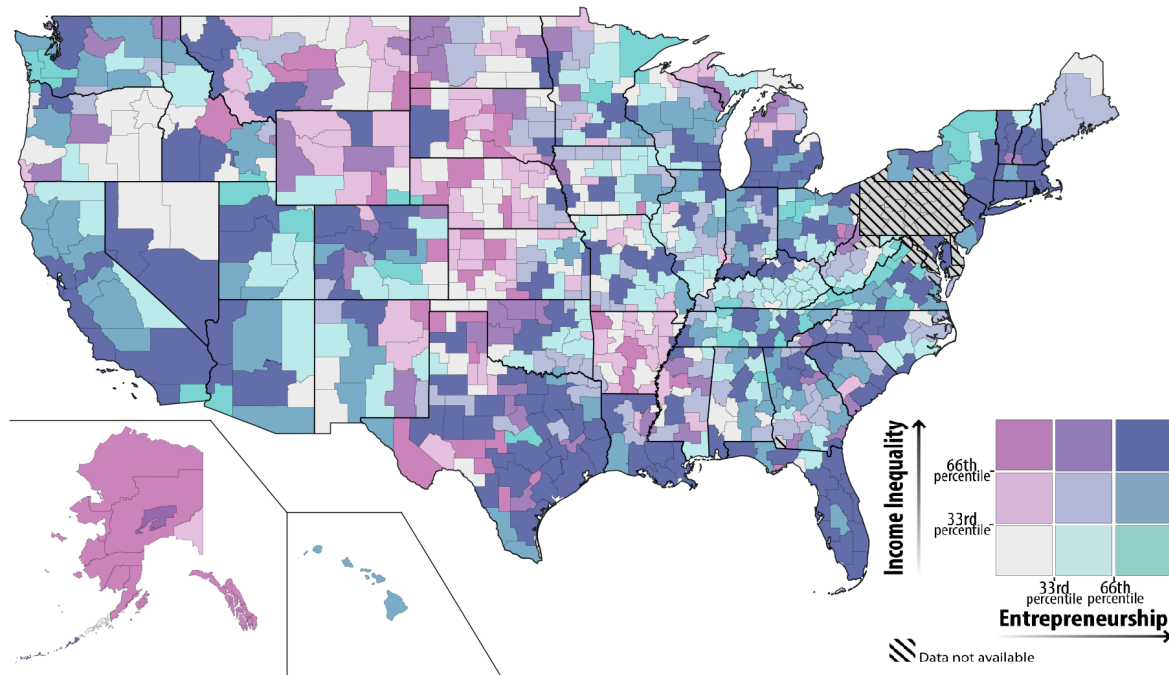
Source: ACS 2005-2015, IRS 1040s 2000-2015, Guzman and Stern (2017). The figure shows the growth of the Gini index (IRS) versus the growth of the RECPI index for the period 1998-2014. Dark blue areas indicate commuting zones where the Gini and RECPI index have increased considerably (both values exceed the 66th percentile of both distribution). White areas are commuting zones where the growth for both variable has not been particularly high (lower than the 33th percentile in both distribution). Pink and cyan areas represents instead commuting zones where the Gini index has increased considerably while the RECPI index has not, and viceversa. DRB approval numbers CBDRB-FY18-164, CBDRB-FY19-342 and CBDRB-FY19-509.
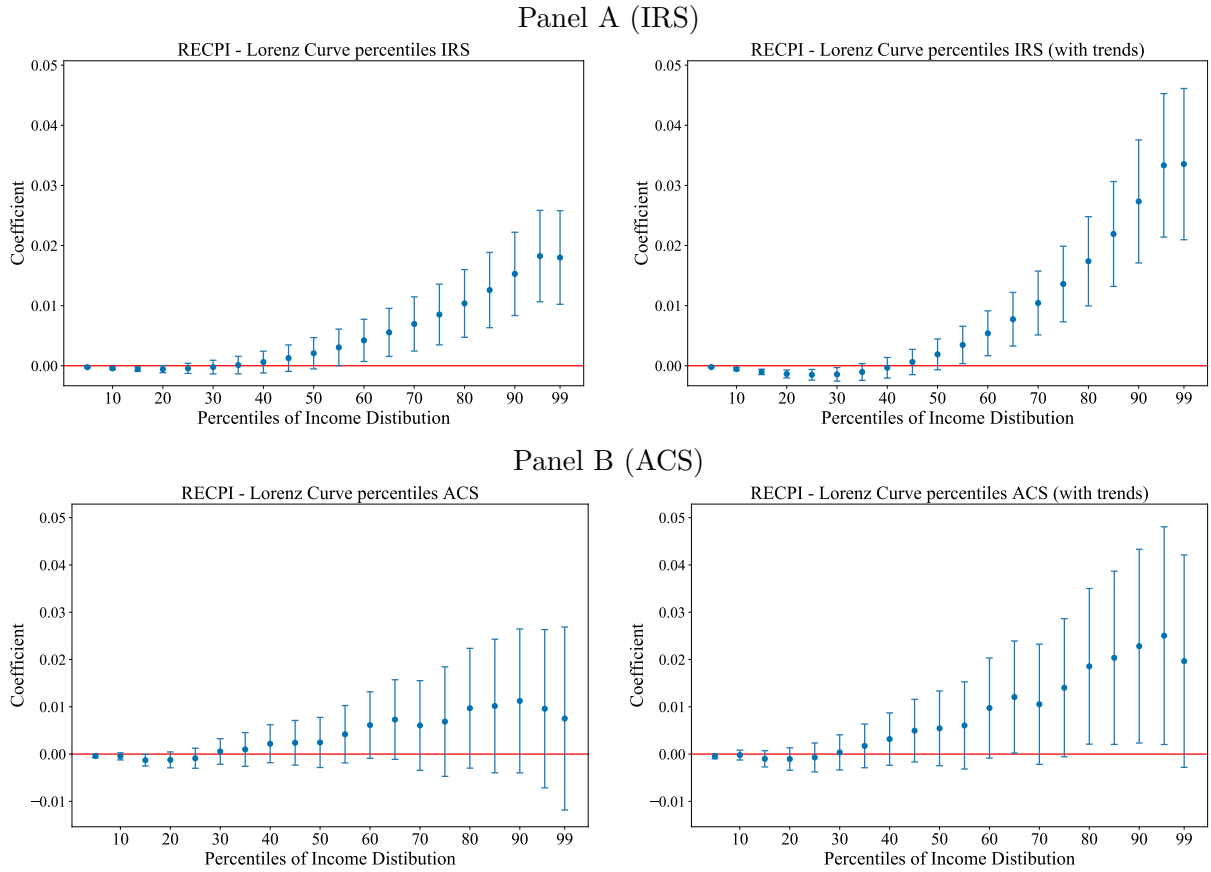
Figure 4: Scatterplot of RECPI and Gini (IRS) for selected years
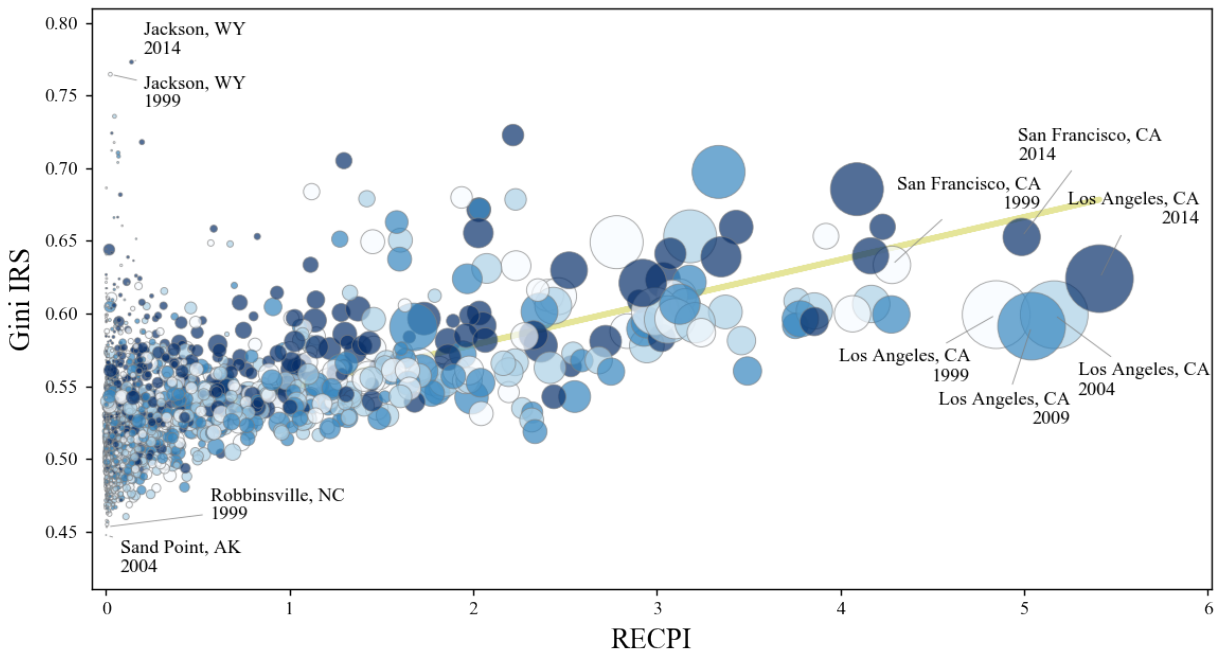
Darker dots are associated with more recent years. For clarity purposes, we consider only the years 1999, 2004, 2009, and 2014. Dot size is proportional to the population of the commuting zone in 2014. DRB approval numbers CBDRB-FY18-164, CBDRB-FY19-342 and CBDRB-FY19-509.

Figure 5: Coefficients of fixed effects regressions of Lorenz Curve percentiles on RECPI
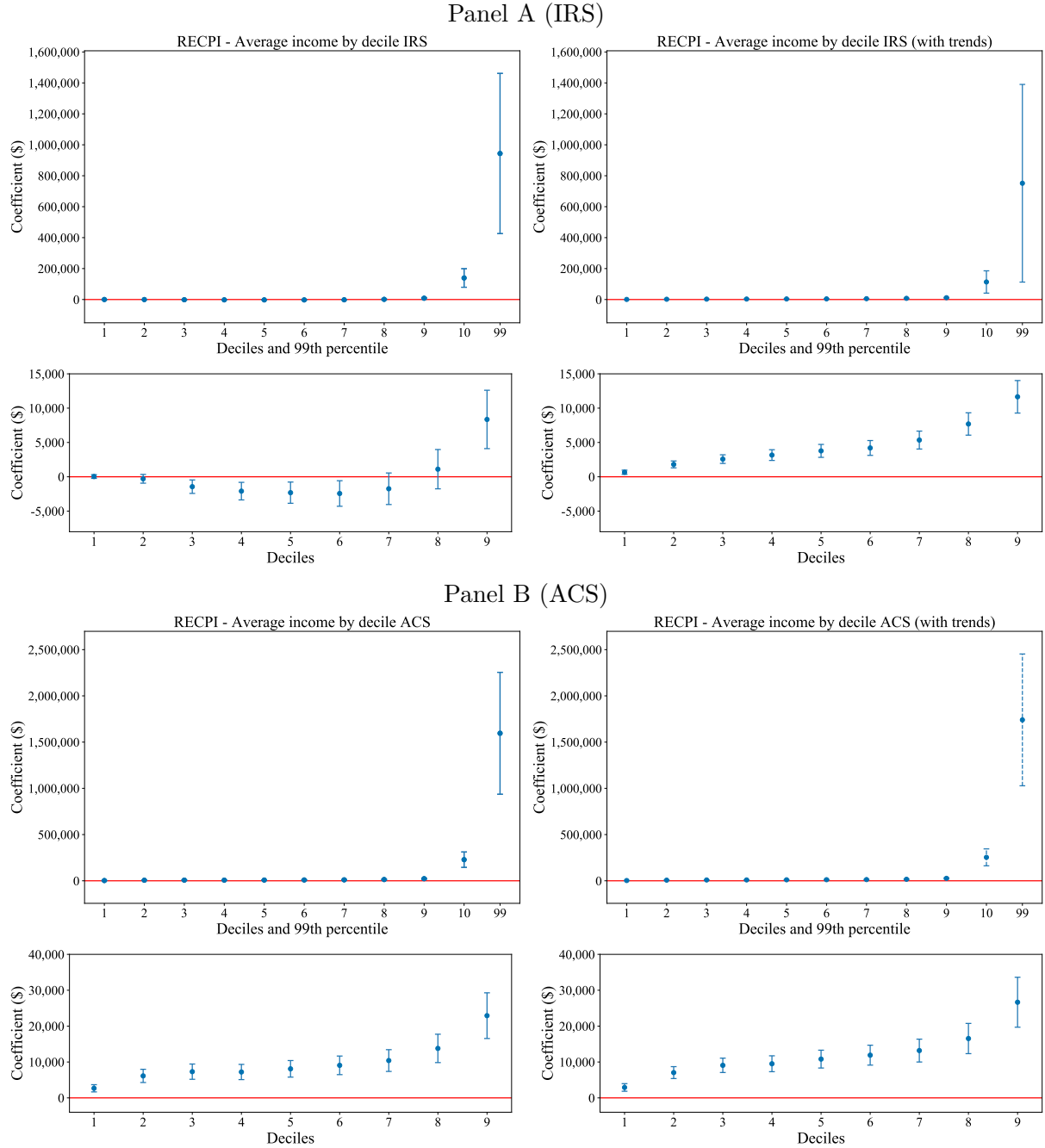
Panel A (IRS)

Panel B (ACS)

Source: ACS 2005-2015, IRS 1040s 2000-2015, Guzman and Stern (2017). The graphs present coefficients of several fixed effects regressions where different Lorenz Curve percentiles are regressed on the RECPI index. All regressions include controls for the ratio of unemployed people on the active population, the ratio of Native American and Alaska Natives, Asian, Black, and Hispanic population on the total population, the ratio of people under 25 and over 65 years of age on total population, the log number of population and per capita personal income in dollars. Standard errors are clustered at the commuting zone level. The length of the vertical segments indicate 10 percent confidence level intervals. Panel A considers IRS-based measures, while Panel B includes ACS-based measures. Both panels include a separate specification with Commuting Zone specific linear trends. DRB approval numbers CBDRB-FY18-164, CBDRB-FY19-342 and CBDRB-FY19-509.

Figure 6: Coefficients of IV regressions of Lorenz Curve percentiles on RECPI
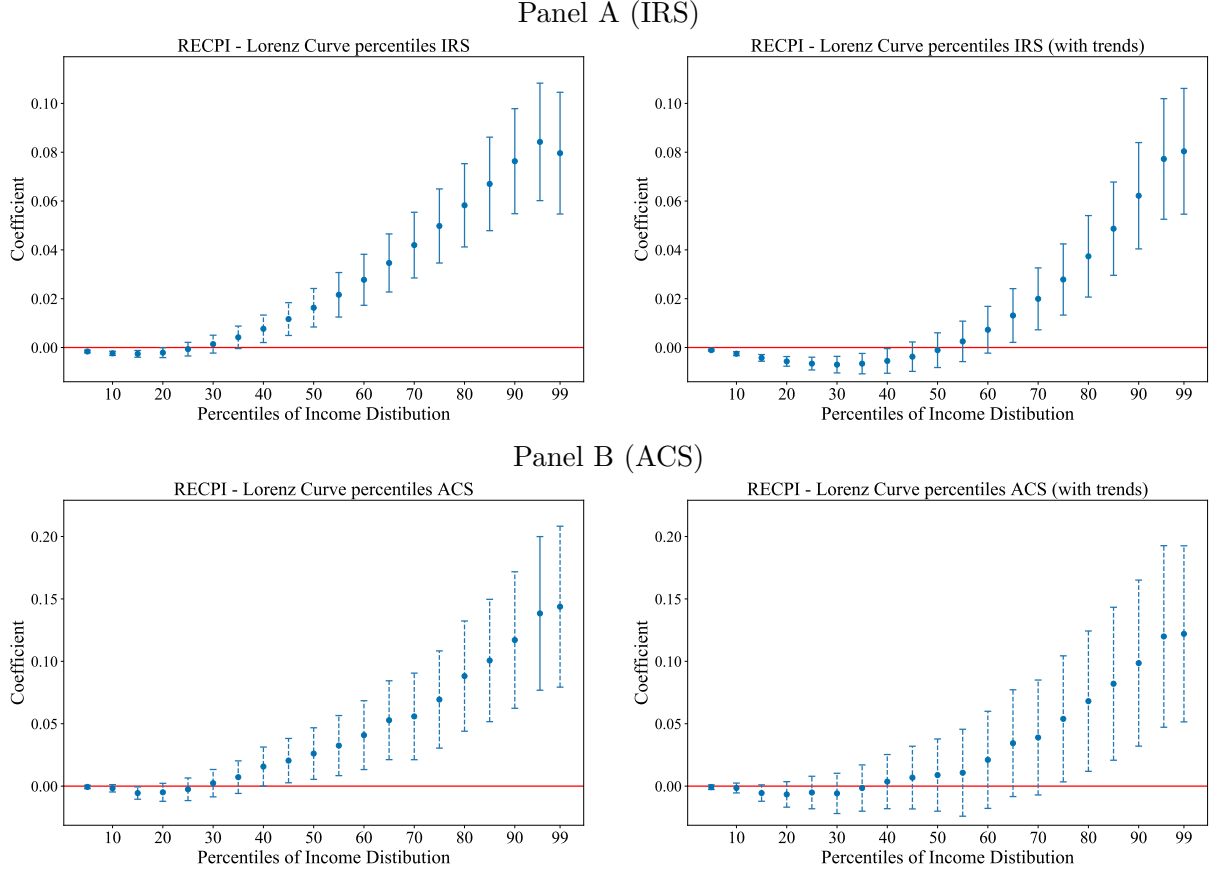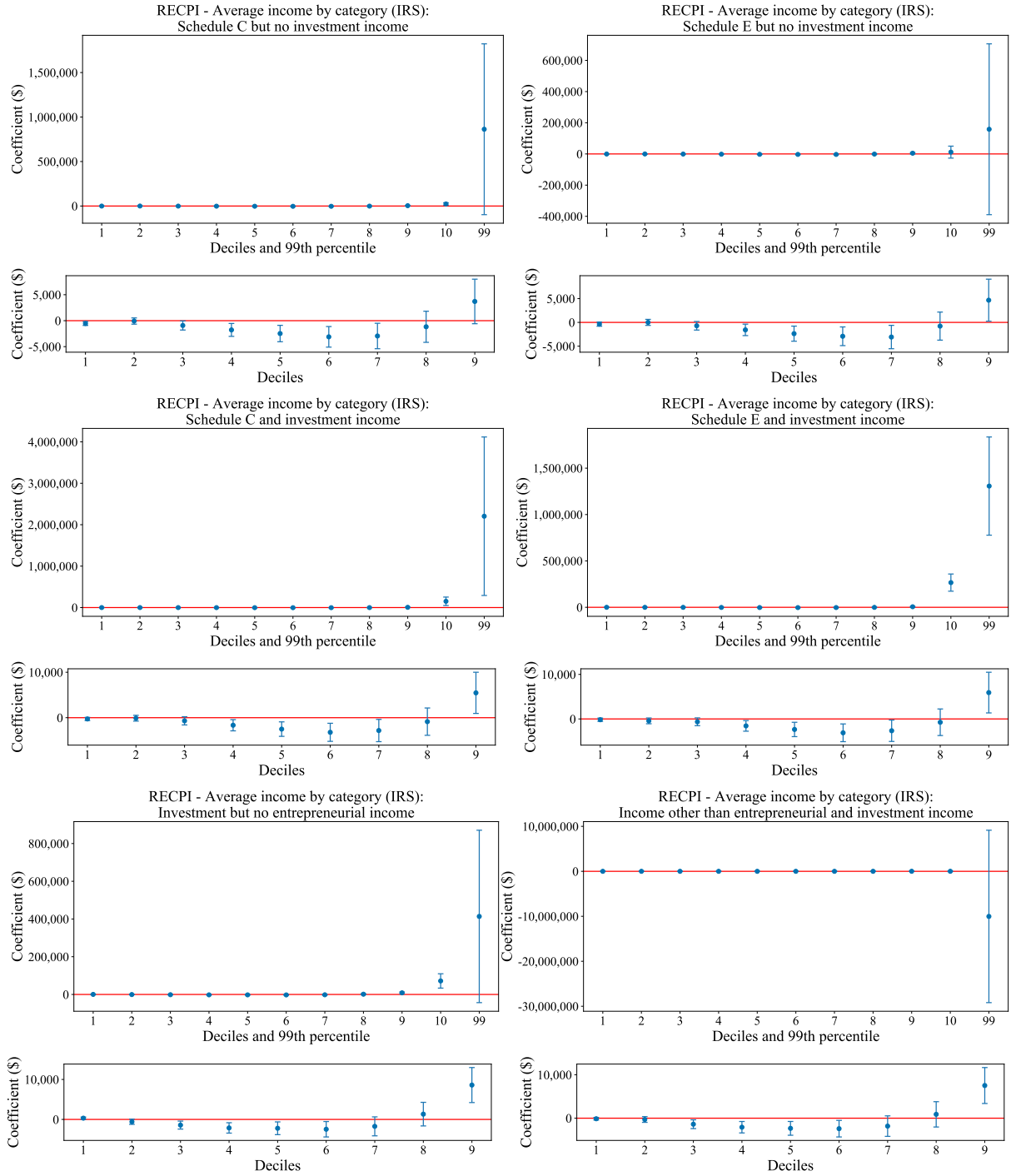
Panel A (IRS)

Panel B (ACS)

Source: ACS 2005-2015, IRS 1040s 2000-2015, Guzman and Stern (2017). The graphs present coefficients of several IV regressions where different Lorenz Curve percentiles are regressed on the RECPI index built from 1988 values of RECPI using a Bartik-style instrument. All regressions include controls for the ratio of unemployed people on the active population, the ratio of Native American and Alaska Natives, Asian, Black, and Hispanic population on the total population, the ratio of people under 25 and over 65 years of age on total population, the log number of population and per capita personal income in dollars. Standard errors are clustered at the commuting zone level. The length of the vertical segments indicate 10 percent confidence level intervals. Panel A considers IRS-based measures, while Panel B includes ACS-based measures. Both panels include a separate specification with Commuting Zone specific linear trends. IV regressions for which the F-statistic of the first stage is less than 10 present dashed standard error lines.

Figure 7: Coefficients of IV regression of average income by deciles and 99th percentile on RECPI
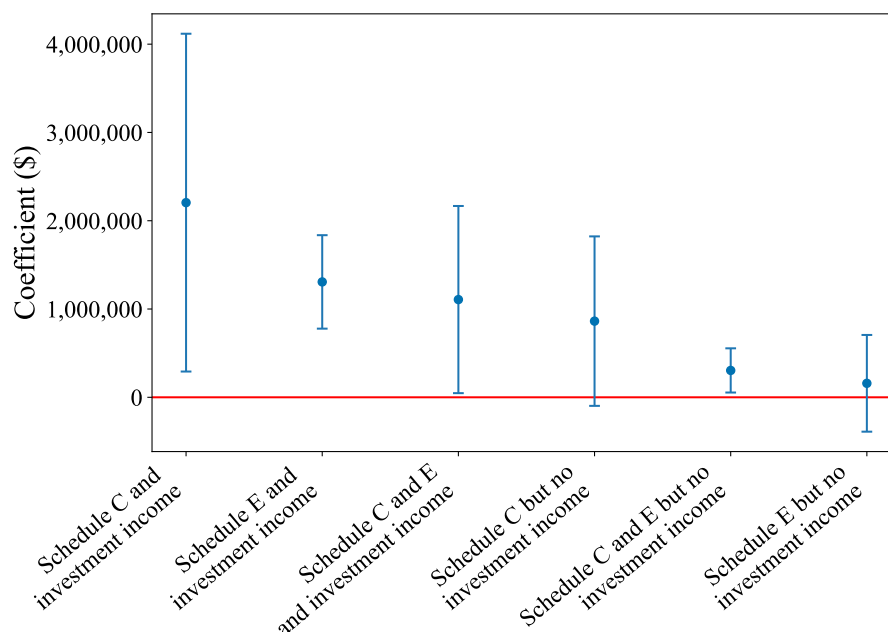
Panel A (IRS)

Panel B (ACS)

Source: ACS 2005-2015, IRS 1040s 2000-2015, Guzman and Stern (2017). The graphs present coefficients of several fixed effects regressions where different levels of average income (which include all the deciles and the top 1 percent of the income distribution) are regressed the RECPI index built from 1988 values of RECPI using a Bartik-style instrument. All regressions include controls for the ratio of unemployed people on the active population, the ratio of Native American and Alaska Natives, Asian, Black, and Hispanic population on the total population, the ratio of people under 25 and over 65 years of age on total population, the log number of population and per capita personal income in dollars. Standard errors are clustered at the commuting zone level. The length of the vertical segments indicate 10 percent confidence level intervals. Panel A considers IRS-based measures, while Panel B includes ACS-based measures. Both panels include a separate specification with Commuting Zone specific linear trends as well as zoomed graphs on the lower section of the income distribution (1st to 9th decile). IV regressions for which the F-statistic of the first stage is less than 10 present dashed standard error lines. DRB approval numbers CBDRB-FY18-164, CBDRB-FY19-342 and CBDRB-FY19-509.

34

Figure 8: Coefficients of IV regression of average income by category on RECPI
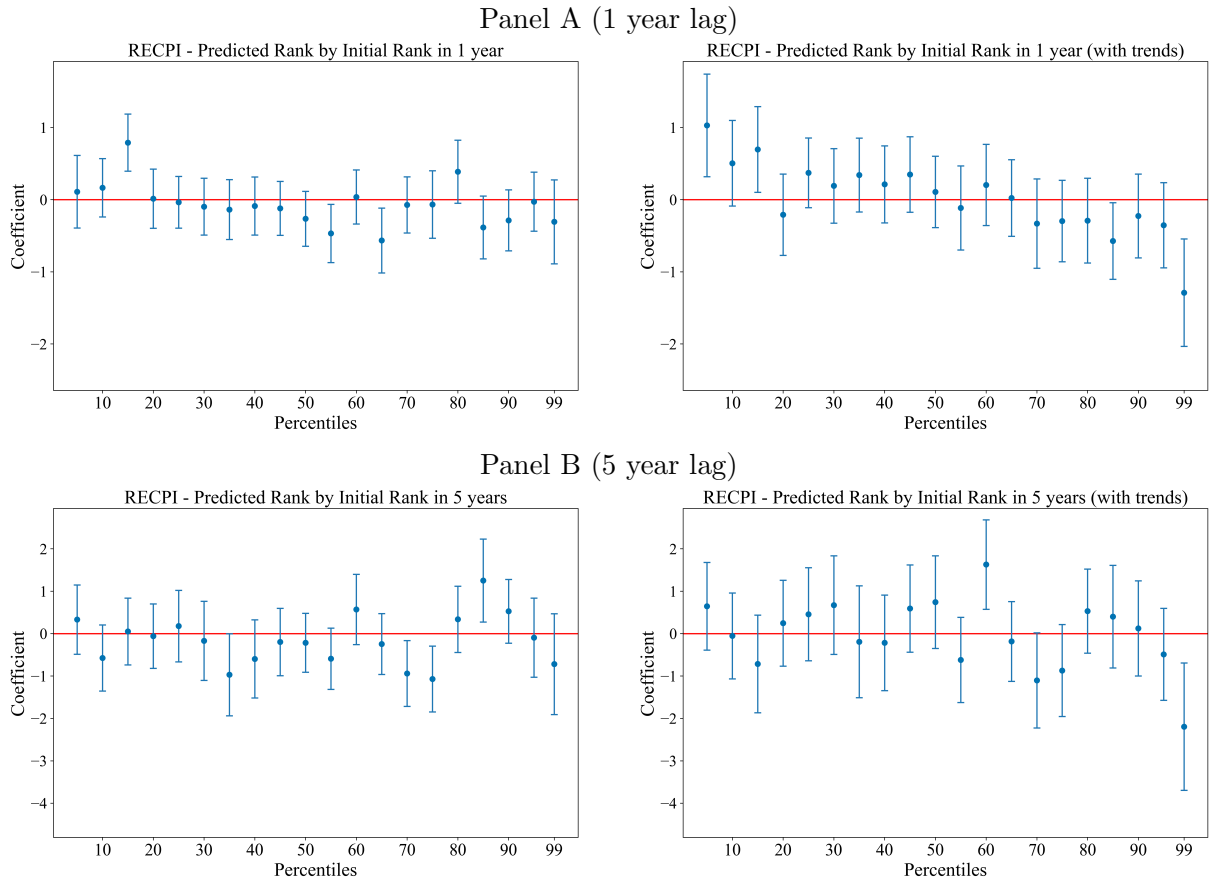

Source: ACS 2005-2015, IRS 1040s 2000-2015, Guzman and Stern (2017). The graphs present coefficients of several fixed effects regressions where different levels of average income split by categories (depending on the type of IRS filings) are regressed the RECPI index built from 1988 values of RECPI using a Bartik-style instrument. All regressions include controls for the ratio of unemployed people on the active population, the ratio of Native American and Alaska Natives, Asian, Black, and Hispanic population on the total population, the ratio of people under 25 and over 65 years of age on total population, the log number of population and per capita personal income in dollars. Standard errors are clustered at the commuting zone level. The length of the vertical segments indicate 10 percent confidence level intervals. IV regressions for which the F-statistic of the first stage is less than 10 present dashed standard error lines. DRB approval numbers CBDRB-FY18-164, CBDRB-FY19-342 and CBDRB-FY19-509.

Figure 9: Coefficients of IV regression of average income by category - Top 1 percentiles on RECPI
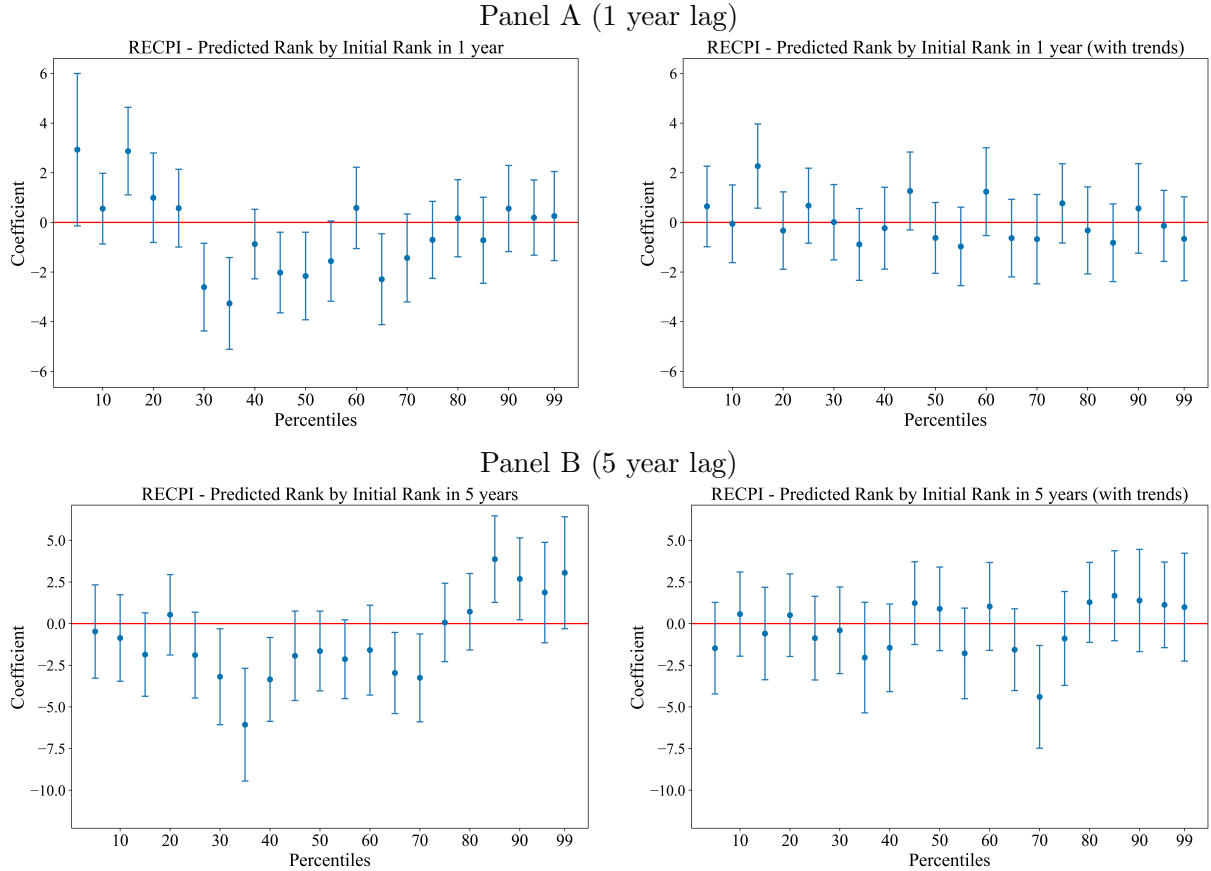

Source: ACS 2005-2015, IRS 1040s 2000-2015, Guzman and Stern (2017). The graph presents coefficients of several fixed effects regressions where different levels of average income split by categories (depending on the type of IRS filings) are regressed the RECPI index built from 1988 values of RECPI using a Bartik-style instrument. The graph only includes coefficients for the top 1 percent in the income distribution for all categories. All regressions include controls for the ratio of unemployed people on the active population, the ratio of Native American and Alaska Natives, Asian, Black, and Hispanic population on the total population, the ratio of people under 25 and over 65 years of age on total population, the log number of population and per capita personal income in dollars. Standard errors are clustered at the commuting zone level. The length of the vertical segments indicate 10 percent confidence level intervals. IV regressions for which the F-statistic of the first stage is less than 10 present dashed standard error lines. DRB approval numbers CBDRB-FY18-164, CBDRB-FY19-342 and CBDRB-FY19-509.

Figure 10: Coefficients of fixed-effects regressions of percentile ranks (1 and 5 years lag) on RECPI

Panel A (1 year lag)

Panel B (5 year lag)

Source: ACS 2005-2015, IRS 1040s 2000-2015, Guzman and Stern (2017). The graphs present coefficients of several fixed effects regressions where different mobility percentiles (measured as percentile ranks with a 1 and 5 years lag) are regressed on the RECPI index. All regressions include controls for the ratio of unemployed people on the active population, the ratio of Native American and Alaska Natives, Asian, Black, and Hispanic population on the total population, the ratio of people under 25 and over 65 years of age on total population, the log number of population and per capita personal income in dollars. Standard errors are clustered at the commuting zone level. The length of the vertical segments indicate 10 percent confidence level intervals. Panel A considers IRS-based measures, while Panel B includes ACS-based measures. Both panels include a separate specification with Commuting Zone specific linear trends. DRB approval numbers CBDRB-FY18-164, CBDRB-FY19-342 and CBDRB-FY19-509.

Figure 11: Coefficients of IV regression of percentile ranks (1 and 5 years lag) on RECPI

Panel A (1 year lag)

Panel B (5 year lag)

Source: ACS 2005-2015, IRS 1040s 2000-2015, Guzman and Stern (2017). The graphs present coefficients of several fixed effects regressions where different mobility percentiles (measured as percentile ranks with a 1 and 5 years lag) are regressed on the RECPI index built from 1988 values of RECPI using a Bartik-style instrument. All regressions include controls for the ratio of unemployed people on the active population, the ratio of Native American and Alaska Natives, Asian, Black, and Hispanic population on the total population, the ratio of people under 25 and over 65 years of age on total population, the log number of population and per capita personal income in dollars. Standard errors are clustered at the commuting zone level. The length of the vertical segments indicate 10 percent confidence level intervals. Panel A considers IRS-based measures, while Panel B includes ACS-based measures. Both panels include a separate specification with Commuting Zone specific linear trends. IV regressions for which the F-statistic of the first stage is less than 10 present dashed standard error lines. DRB approval numbers CBDRB-FY18-164, CBDRB-FY19-342 and CBDRB-FY19-509.
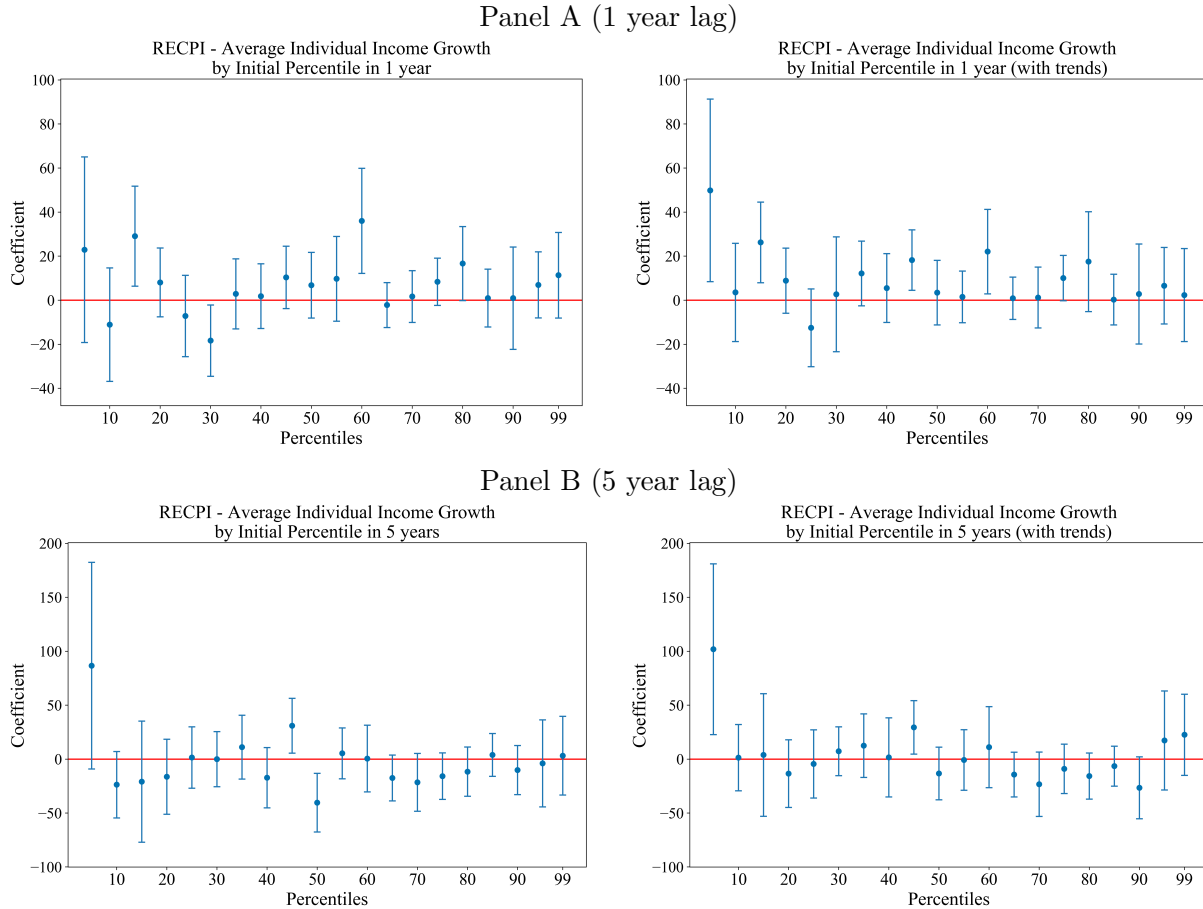
10 Appendix

10.1 Additional mobility measures

As a further robustness check for our mobility-related results we also examine additional mobility concepts. We consider the effect of entrepreneurship on income mobility profiles – the average individual income growth rate between the base and final year (1 and 5 years are considered) for each percentile of the base year income distribution. A1 present the results for our usual specifications. One and five year models not including trends do not suggest any clear relationship between our variables of interest. However, models with trends seem to indicate the presence of a positive, significant relationship between entrepreneurship and average individual income of people belonging to the 5th percentile of the distribution. This result appears to be driven by business losses – individuals with large business losses in the base year will end up in the bottom vigintile of the income distribution. However, it appears that these individuals earn much higher incomes 1 and 5 years later in CZs with large entrepreneurship shocks. This is consistent with our results in section 6, showing that the benefits of entrepreneurship accrue largely to individuals with business income.

10.2 Appendix figures and tables

Figure A1: Coefficients of IV regressions of average individual income growth by initial percentile (1 and 5 years lag) on RECPI

Panel A (1 year lag)

Panel B (5 year lag)

Source: ACS 2005-2015, IRS 1040s 2000-2015, Guzman and Stern (2017). The graphs present coefficients of several fixed effects regressions where different mobility percentiles (measured as percentile ranks with a 1 and 5 years lag) are regressed on the RECPI index built from 1988 values of RECPI using a Bartik-style instrument. All regressions include controls for the ratio of unemployed people on the active population, the ratio of Native American and Alaska Natives, Asian, Black, and Hispanic population on the total population, the ratio of people under 25 and over 65 years of age on total population, the log number of population and per capita personal income in dollars. Standard errors are clustered at the commuting zone level. The length of the vertical segments indicate 10 percent confidence level intervals. Panel A considers IRS-based measures, while Panel B includes ACS-based measures. Both panels include a separate specification with Commuting Zone specific linear trends. IV regressions for which the F-statistic of the first stage is less than 10 present dashed standard error lines. DRB approval numbers CBDRB-FY18-164, CBDRB-FY19-342 and CBDRB-FY19-509.

References

Acs, Z. J. and Audretsch, D. B. (1988). Innovation in large and small firms: an empirical analysis. *The American economic review*, pages 678–690.

Aghion, P., Akcigit, U., Bergeaud, A., Blundell, R., and Hémous, D. (2015). Innovation and top income inequality. *National Bureau of Economic Research*.

Aghion, P. and Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, 60:323–351.

Autor, D. H., Manning, A., and Smith, C. L. (2016). The contribution of the minimum wage to us wage inequality over three decades: A reassessment. *American Economic Journal: Applied Economics*, 8(1):58–99.

Bartik, T. J. (1991). Who benefits from state and local economic development policies?

Berkes, E. and Gaetani, R. (2019). Income segregation and rise of the knowledge economy. *Available at SSRN 3423136*.

Binder, D. and Kovacevic, M. (1995). Estimating some measures of income inequality from survey data: An application of the estimating equations approach. *Survey Methodology*, 21.

Card, D. and DiNardo, J. E. (2002). Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles. *Journal of Labor Economics*, 20(4):733–783.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics*, 129(4):1553–1623.

Cobb, J. A. and Stevens, F. G. (2017). These unequal states: Corporate organization and income inequality in the united states. *Administrative Science Quarterly*, 62(2):304–340.

Cohen, W. M., Nelson, R. R., and Walsh, J. P. (2000). Protecting their intellectual assets: Appropriability conditions and why us manufacturing firms patent (or not). Technical report, National Bureau of Economic Research.

Darrow, J. E. (1989). Tax considerations—from the company's standpoint—in structuring venture capital investments. *The Business Lawyer*, pages 233–260.

Davis, S. J. and Haltiwanger, J. (1992). Gross job creation, gross job destruction, and employment reallocation. *The Quarterly Journal of Economics*, 107(3):819–863.

Decker, R., Haltiwanger, J., Jarmin, R., and Miranda, J. (2014). The role of entrepreneurship in us job creation and economic dynamism. *Journal of Economic Perspectives*, 28(3):3–24.

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica*, 64(5):1001–1044.

Frank, R. H. and Cook, P. (1996). *The winner-take-all society: Why the few at the top get so much more than the rest of us.*

Goldin, C. D. and Katz, L. F. (2009). *The race between education and technology.* Harvard University Press.

Guzman, J. and Stern, S. (2017). Nowcasting and placecasting entrepreneurial quality and performance. *NBER/CRIW Conference Volume on Measuring Entrepreneurship.*

Halvarsson, D., Korpi, M., and Wennberg, K. (2018). Entrepreneurship and income inequality. *Journal of Economic Behavior & Organization*, 145:275–293.

Hémous, D. and Olsen, M. (2014). The Rise of the Machines: Automation, Horizontal Innovation and Income Inequality.

Hurst, E. and Pugsley, B. W. (2011). What Do Small Businesses Do? *NBER Working Paper.*

Katz, L. F. and Murphy, K. M. (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics*, 107(1):35–78.

Keister, L. A. and Moller, S. (2000). Wealth inequality in the United States. *Annual Review of Sociology*, 26(1):63–81.

Marshall, A. (1920). Principles of economics. *London: Mac-Millan*, pages 1–627.

Paunov, C. and Guellec, D. (2017). Digital innovation and the distribution of income. *Measuring and accounting for innovation in the 21st century, NBER, Book Series in Income and Wealth.*

Piketty, T. and Goldhammer, A. (2014). *Capital in the Twenty-first Century.* The Belknap Press of Harvard University Press.

Piketty, T. and Saez, E. (2003). Income inequality in the united states, 1913-1998. *Quarterly Journal of Economics*, 118(1).

Rosen, S. (1981). The economics of superstars. *The American economic review*, 71(5):845–858.

Saxenian, A. (1994). Regional networks: industrial adaptation in silicon valley and route 128.

Schoar, A. (2010). The Divide between Subsistence and Transformational Entrepreneurship. *Innovation Policy and the Economy*, 10(1):57–81.

Schumpeter, J. (1942). Creative destruction. *Capitalism, socialism and democracy*, 825.

Smith, M., Yagan, D., Zidar, O. M., and Zwick, E. (2019). Capitalists in the twenty-first century.

Sørensen, J. B. (2007). Organizational diversity, labor markets, and wage inequality. *American Behavioral Scientist*, 50(5):659–676.

Sorensen, J. B. and Sorenson, O. (2007). Corporate Demography and Income Inequality. *American Sociological Review*, 72(5):766–783.

Van Kerm, P. (2009). Income mobility profiles. *Economics Letters*, 102(2):93–95.